A Neural Network Approach for Chinese Character Recognition

Sheng-De Wang and Chung-Chi Pan

Department of Electrical Engineering
National Taiwan University, Taipei 10764, Taiwan

Abstract—A Neural network model adapted from Fukushima's Neocognitron is applied to the pattern recognition of Chinese characters. Chinese characters are well-known for their non-alphabetic and two-dimensional features. In this paper, Chinese characters are viewed as two-dimensional patterns and some simple sub-patterns, called primitives, are identified for hierarchical processing in a neural network. Simulation results show the feasibility and the effectiveness of this approach.

1. Introduction

Human brain, which can learn, think, and memorize, is perhaps the most complex mechanism that can hardly be understood. After years of investigation, neural networks have turned out to be one promising approach for the cognitive science. Many research works have been done on the field of applying neural network to computer vision, image processing and pattern recognition. Fukushima's Neocognitron [1]-[4] have been shown to be powerful for recognizing handwritten Arabic numerals. The ADALINE of Widrow et al. [5] also aimed to modeling the capability of retina and brain image processing. Grossberg and his colleagues [6][7] have developed a class of neural network for self-organizing pattern recognition.

One of benefits that the neural network approach for pattern recognition surpasses the conventional approach is its ability to cope with shift in position and distortions in shape of the stimulus patterns. Neocognitron is capable of recognizing stimulus patterns correctly even if the patterns are shifted in positions or distorted in shape.

A neural network model proposed by Malsburg[8], based on the concept of labeled graph matching, also solves the position-invariant problem in pattern recognition. Recently, Menon[9] has successfully applied the Neocognitron to recognize and classify three complex images. In this paper, we try to explore the potential use of the pattern feature extraction and selective attention capability of Neocognitron for Chinese character recognition. Since the patterns of Chinese characters are more complex than that of Arabic numerals, some cares must be taken when the Neocognitron is applied for recognizing Chinese characters.

2. The Network Architecture

The proposed network for Chinese character recognition is adapted from the Neocognitron proposed by Fukushima, et al. [4]. The Neocognitron is a multilayer network consisting of a cascade of layers of neuron-like cells. Each layer is composed of two stages of cells: stage U_S consisting of S cells followed by stage U_C consisting of C cells. These cells are analog type; that is, their inputs and outputs are non-negative analog values. Each cell has two kind of input stimuli- excitatory and inhibitory stimuli. Fig. 1 shows the basic structure of one neuron. Each C cells has adjustable input synapses, which can be reinforced through learning, leading from a group of C cells and selectively responds to a specific stimulus pattern. This group of C cells is sometimes called the receptive field of the S cell. Each C cell has fixed input synapses leading from a group of S cells which have similar pattern in their receptive field. This means that the C cell will selectively respond to a specific pattern presented at the input layer which activates at least one presynaptic S cell. This property makes the network respond to the same stimulus in different positions.

In the whole network, as shown in Fig. 2, stage U_S and stage U_C are arranged alternately. The stages U_S and U_C in lth layer are denoted by U_{Sl} and U_{Cl}, respectively. The U_{S1} is preceded by an input layer U_{C0} to which the pattern is presented. The output layer U_{C4} will respond selectively to the input pattern. Cells in each layer are arranged into several planes. Each plane consists of a number of cells and is trained to recognize one primitive. A primitive is a simple pattern which is part of a pattern. For a set of patterns, we can identify primitives which can be combined to constitute various patterns. The more deeper layer a plane belongs, the more complex primitive it responds for. Hence, the C cells in the deepest layer respond selectively to the whole pattern presented at input layer. For the purpose of shift invariant, cells in the same plane have the same set of input synapses, so that they could respond to the same pattern.

The number of planes in each layer could be different, depending on the number of primitives that all the training patterns have. Besides, each layer contains subsidiary inhibitory cells, called V-cells. There are two kind of inhibitory cells : V_S-cell and V_C-cell, which are located in U_S layer and U_C layer, respectively. Fig. 4 shows the connections between excitatory cells and inhibitory cells. If the incoming stimuli are large enough, excitatory cell will issue a nonnegative excitatory stimulus. Unlike the excitatory cell, if there are any responses in the receptive field, inhibitory cell will output the stimulus directly. There is no limitation of layers used in Neocognitron. Fukushima[4] suggests that for complex patterns it is better to use more layers in Neocognitron.

Now, we will describe the outputs of the cells in the network with numerical expressions. The output of a U_S-cell located at $\mathbf{n}$ in the k-th plane of the lth layer is given by

$$u_{sl}(k_l,\mathbf{n}) = r_l * \varphi\left[\frac{1+\sum_{k_{l-1}=1}^{K_{l-1}} \sum_{\upsilon\in S_l} a_l(k_{l-1}, \upsilon, k_l) * u_{cl-1}(k_{l-1}, \mathbf{n}+\upsilon)}{1+\frac{r_l}{1+r_l} * b_l(k_l) * v_{cl-1}(\mathbf{n})} - 1\right] \tag{1}$$

where $\mathbf{n}$ denotes a two-dimensional coordinates, K_{l-1} is the number of planes in $(l-1)$th layer, S_l is the receptive field, and r_l is a parameter which control the intensity of the inhibitory stimulus to the U_S-cell. In the case of $l=1$, $u_{cl-1}(k_{l-1},\mathbf{n})$ represents $u_0(\mathbf{n})$. Here, $a_l(k_{l-1}, \upsilon, k_l)$ and $b_l(k_l)$ represent the values of excitatory stimulus and inhibitory stimulus, respectively. $\varphi[.]$ is a nonnegative function which is defined as follows:

$$\varphi[x] = \begin{cases} x & (x \geqq 0) \\ 0 & (x < 0) \end{cases} \tag{2}$$

As described in (1), the U_S-cell receives signals from all the planes in the previous layer. Since each plane is used to recognize one primitive and all the patterns are composed of these primitives, informations must be collected from all these planes.

The inhibitory cell $v_{cl-1}(\mathbf{n})$ receives stimuli from U_{cl-1} layer and its response is given by

$$v_{cl-1}(\mathbf{n}) = \sqrt{\sum_{k_{l-1}=1}^{K_{l-1}} \sum_{\upsilon\in S_l} c_{l-1}(\upsilon) * u_{cl-1}^2(k_{l-1}, \mathbf{n}+\upsilon)} \tag{3}$$

where $c_{l-1}(u)$ is a fixed weight and is a monotonically decreasing function of $|\upsilon|$:

$$c_l(\upsilon) = \frac{1}{C(l)} \alpha_l^{|\upsilon|} \tag{4}$$

where α_l is a small positive constant(<1) which determines how quickly the weights fall when $|\upsilon|$ is increase, and $C(l)$ is a normalized constant

$$C(l) = \sum_{k_l=1}^{K_l} \sum_{\upsilon\in S_l} \alpha_l^{|\upsilon|} \tag{5}$$

The output of U_C-cell of the k-th plane in the l-th layer is given by

$$u_{cl}(k_l, \mathbf{n}) = \psi\left[\frac{1 + \sum_{\upsilon \in D_l} d_l(\upsilon) * u_{sl}(k_l, \mathbf{n} + \upsilon)}{1 + v_{sl}(\mathbf{n})} - 1\right] \tag{6}$$

where

$$\psi[x] = \begin{cases} \dfrac{x}{\alpha + x} & (x \geqq 0) \\ 0 & (x < 0) \end{cases} \tag{7}$$

It can be seen from (6) that U_C-cell only receives stimuli from a receptive field D_l on one plane in the preceded layer. The input-output function $\psi[x]$ limits the response of U_C-cell within [0,1]. The inhibitory signal in (6) is generated from U_{Sl}-cell as follows.

$$v_{sl}(\mathbf{n}) = \frac{1}{K_l} \sum_{k_l=1}^{K_l} \sum_{\upsilon \in D_l} d_l(\upsilon) * u_{sl}(k_l, \mathbf{n} + \upsilon) \tag{8}$$

The value of $d_l(\upsilon)$ is defined as in the case of $c_l(\upsilon)$ so as to decrease monotonically with respect to $|\upsilon|$ and is also an unmodifiable connection.

3. Training the Network

3.1 Synapse reinforcement phase

Before using Neocognitron for pattern classification, we should train it with patterns. The synaptic connections of Neocognitron are reinforced by means of a supervised learning scheme, that is, a learning-with-a-teacher process. Training is done incrementally from left layers to right layers. To train the leftmost layer, the most basic primitives are presented at input layer. As the training proceeds, the complexity of primitive pattern are increased. When we present one primitive to input plane, Neocognitron automatically choose a plane to represent this primitive by using the concept of the seed cell. A cell with maximum response in a plane is selected as the seed cell of the plane. Synaptic connections of the cells are reinforced depending on the values of stimuli given to these synaptic connections. That is, we just reinforce those synaptic connections through which nonzero incoming signals. Suppose $u_{Sl}(k_l, \mathbf{n_s})$ is the seed cell. The adjustable synapses $a_l(k_l, \upsilon, \mathbf{n_s})$ and $b_l(k_l)$ are reinforced by the amount.

$$\Delta a_l(k_{l-1}, \upsilon, k_l) = q_l * c_{l-1}(\upsilon) * u_{cl-1}(k_{l-1}, \mathbf{n_s} + \upsilon) \tag{9}$$

$$\Delta b_l(k_l) = q_l * v_{cl-1}(\mathbf{n_s}) \tag{10}$$

where k_{l-1} represents the plane number in U_{Cl-1} layer. and q_l is a positive constant which controls the amount of reinforcement. After training, the input synapses of the seed cell are duplicated to all cells in the plane. Note that in (9) and (10) the ceed cell coordinate $\mathbf{n_s}$ is dropped for left-hand side since synapses $a_l(k_l, \upsilon, \mathbf{n_s})$ and $b_l(k_l)$ are the same for all cells in the same plane.

3.2 Output feedback retraining phase

In the synapse reinforcement phase, the synaptic connections of S cells are adjusted for individual primitive or pattern. For the purpose of correctly recognizing Chinese characters, the patterns to be classified, other than their primitive patterns, should be restrained again. We call this phase the output feedback retraining phase. In this case, we present the training patterns used for U_{S4} layer again, one at a time. When presenting that training pattern, there are some responses in the output cells. If there are ambiguous responses in the output cells, an update of synaptic strength of U_{S4} should be fired as shown in Fig. 4. The amount of reinforcement is given as (9) with an appropriate value of q_l. In this retraining process, we just increase the excitatory connections of the cells in that plane. This step is continued until the desired cell in output layer has the maximum response value.

3.3 Calibration of inhibitory signal

For Chinese characters, strokes are often intersected and hence has a cross point between them. For example, the sub-pattern "+" has a cross point between the two strokes "—" and "|". All these three sub-patterns are primitives that should be trained for recognizing in a plane. Unfortunately, there is a phenomenon that will degrade the recognizing process. Fig. 5 shows the situation that the sub-pattern "+" presented in input layer will deactivate the response of the of seed cell in the " —" or "|" recognizing plane since the inhibitory signal coming from the pattern "+" exceeds the excitatory signal coming from " —" or "|" in this particular seed cell. One approach to alleviate this defect is to readjust the strength of synapses $b_l(k_l)$ such that the synaptic strength is temporally decreased an amount proportional to inhibitory stimulus.

$$\Delta b_l(k_l) = -\delta\, v_{cl-1}(n) \tag{11}$$

where δ is a positive constant.

4. Experiments

A software program for simulating the Neocognitron has been constructed at a SUN/4 computer, and several experiments have also been carried out for recognizing different sets of Chinese characters. In the first experiment, we successfully use the Neocognitron without the retraining process for recognizing 6 Chinese characters. However, when the number of characters is increased to 10 or more characters, Neocognitron failed to perfectly classify them. we then try another experiment by making use of the output feedback scheme and the inhibitory calibration process, and find that Neocognitron can correctly recognize up to 50 characters. Due to the limitation of simulation time and memory space, character set more than 50 elements is not allowed in the present research. It is further shown that with output feedback retraining and inhibitory calibration process the Neocognitron is less sensitive to the free parameter r_l in (1) . r_l is a parameter governing the intensity of the inhibitory stimulus to the Us-cell. In an experiment for recognizing 20 characters, several sets of parameter r_l are found to be suitable, as shown in Table 1. Table 1 also shows the other parameters used in the experiment. For this 20-character experiment, training primitives used for layer 1, 3 and 4 are shown in Fig. 6, 7, and 8, respectively. We use 43 primitive patterns for training layer 2 , which are not shown here due to the page limit. In these experiments, we find that the network has very nice property of modularity. By modularity we mean that the network can easily be expanded to recognize more characters by just adding and training some extra primitives needed to constitute new characters. The previously trained primitives are still useful and valid for larger character set. Once all the primitives of characters are trained for recognizing in hidden layers, each time when we would like to add new characters as target pattern, what we need to do is to extend the output layers and activate an output feedback retraining process.

5. Conclusions

This paper have demonstrated the potential use of neural network approach for Chinese character recognition. Neocognitron enhanced with the retraining process and the inhibitory calibration is shown to be powerful for recognizing Chinese characters. Although not shown here, our experiments also show that the approach has good noise tolerance property. Together with the properties of shift-invariant and modularity, the Neocognitron have provided an alternative for solving difficult pattern recognition problems—such as Chinese character recognition.

References

[1] K. Fukushima and Sei Miyake. "Neocognitron: A new algorithm for pattern recognition

tolerance of deformations and shifts in position". *Pattern Recognition* , vol. 15, pp 455 - 469, 1982.
[2] K. Fukushima, S. Miyake, and T. Ito. "Neocognitron: a neural network model for a mechanism of pattern recognition". *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13, pp. 826-834, 1983.
[3] K. Fukushima. "A neural network model for selective attention in visual pattern recognition". *Biological Cybernetics*, Vol. 55(1), pp. 5-15, 1986.
[4] K. Fukushima. "Neocognitron: A hierarchical neural network capable of visual pattern recognition". *Neural Networks*, Vol 1, pp. 119-130, 1988.
[5] B. Widrow. "Neural Nets for Adaptive Filtering and Adaptive Pattern Recognition". *IEEE Computer*, pp. 25 - 39, Mar, 1988.
[6] G. A. Carpenter and S. Grossberg. "A massively parallel architecture for a self-organizing neural pattern recognition machine". *Computer vision, Graphics, and image processing* , Vol.37, 54 -115, 1987.
[7] G. A. Carpenter and Stephen Grossberg. "ART 2: self-organization of stable category recognition codes for analog input patterns". *Applied optics*, Vol. 26, No. 23, pp. 4919 - 4930, Dec. 1987.
[8] C. von der Malsburg. "Self-organization of orientation sensitive cells in the striate cortex." *Kybernetik*, Vol. 14, pp. 85-100, 1973.
[9] M. M. Menon and K.G.Heinemann. "Classification of patterns using a self-organizing neural network". *Neural networks*, Vol 1, pp 201-215, 1988.

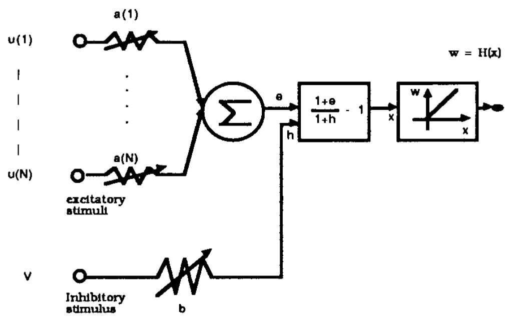

Fig. 1. Basic structure of one neuron

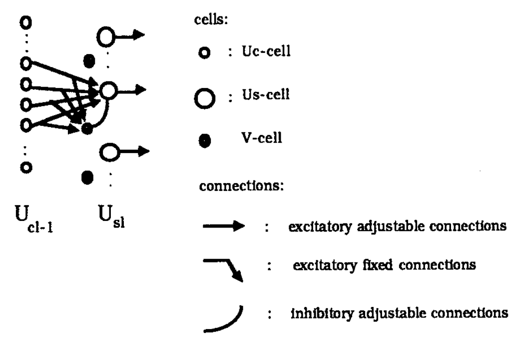

Fig. 3. Synaptic connections between excitatory cells and inhibitory cells.

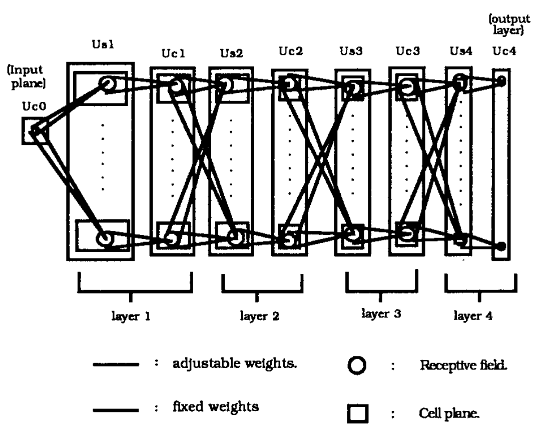

Fig. 2. The architecture of the Neocognitron.

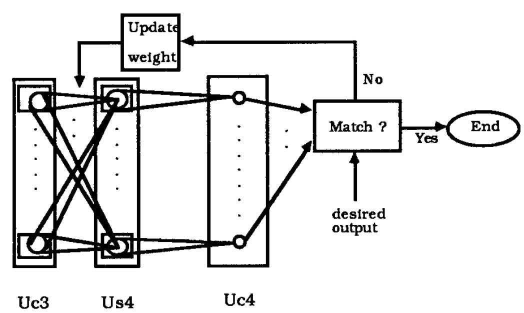

Fig. 4. Retraining the synaptic connections by output feedbaack.

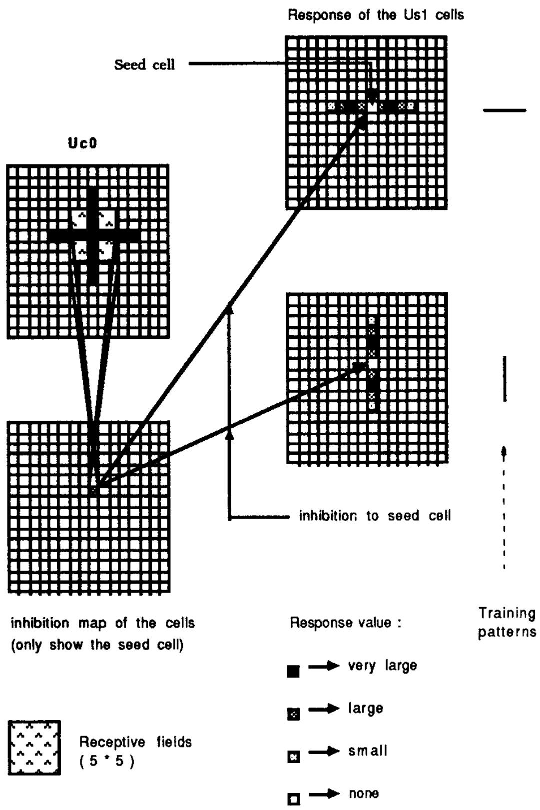

Fig. 5. Example of undesired response of seed cell.

Level	No. planes	S-plane size	C-plane size	Receptive field S-layer	Receptive field C-layer	q_l
1	13	16*16	12*12	5*5	5*5	60000
2	43	12*12	8*8	5*5	5*5	60000
3	32	8*8	8*8	5*5	5*5	60000
4	20	4*4	1*1	5*5	4*4	6000

experiments / r_l	1	2	3	4	5	6	7	8	9	10
r_1	5	4.9	4.8	4.7	4.6	4.5	4.4	4.3	4.2	4.1
r_2	4	3.9	3.8	3.7	3.6	3.5	3.4	3.3	3.2	3.1
r_3	3	2.9	2.8	2.7	2.6	2.5	2.4	2.3	2.2	2.1
r_4	2	1.9	1.8	1.7	1.6	1.5	1.4	1.3	1.2	1.1

Table 1. Parameters used for recognizing 20 Chinese characters.

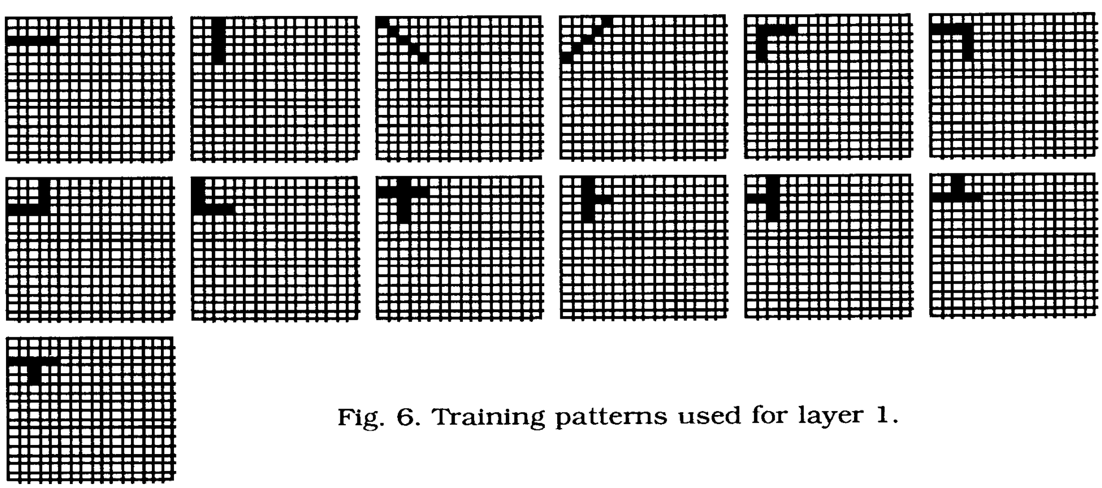

Fig. 6. Training patterns used for layer 1.

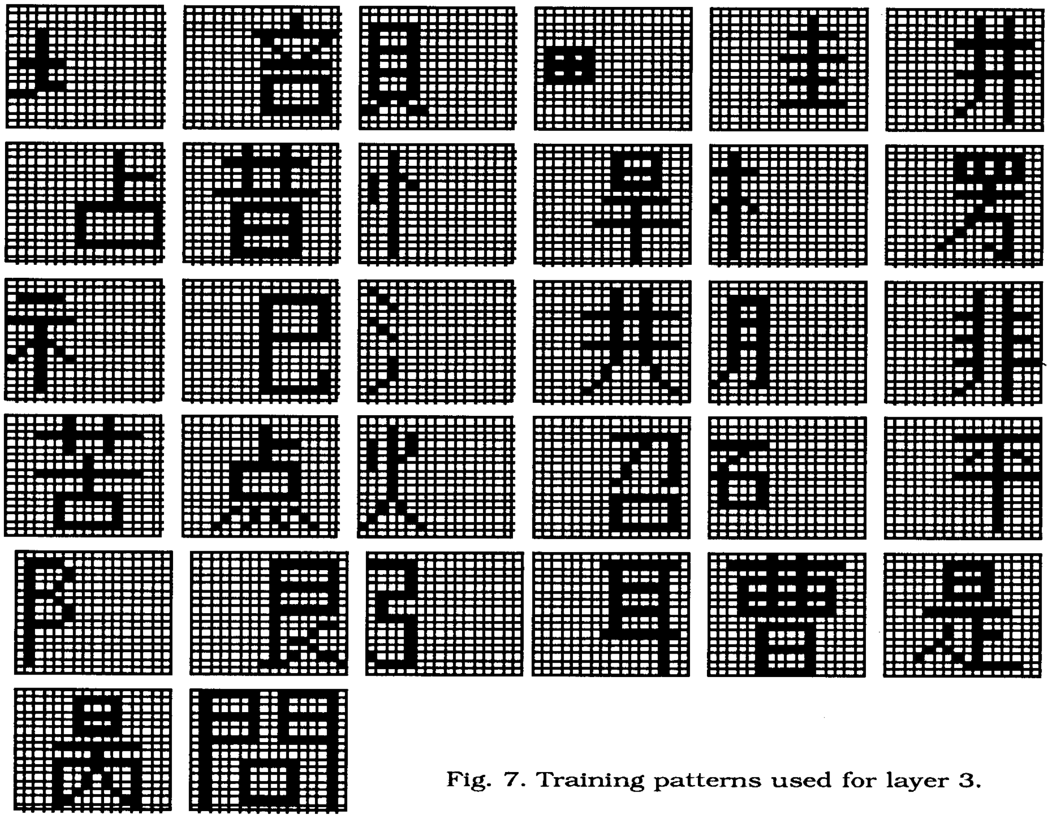

Fig. 7. Training patterns used for layer 3.

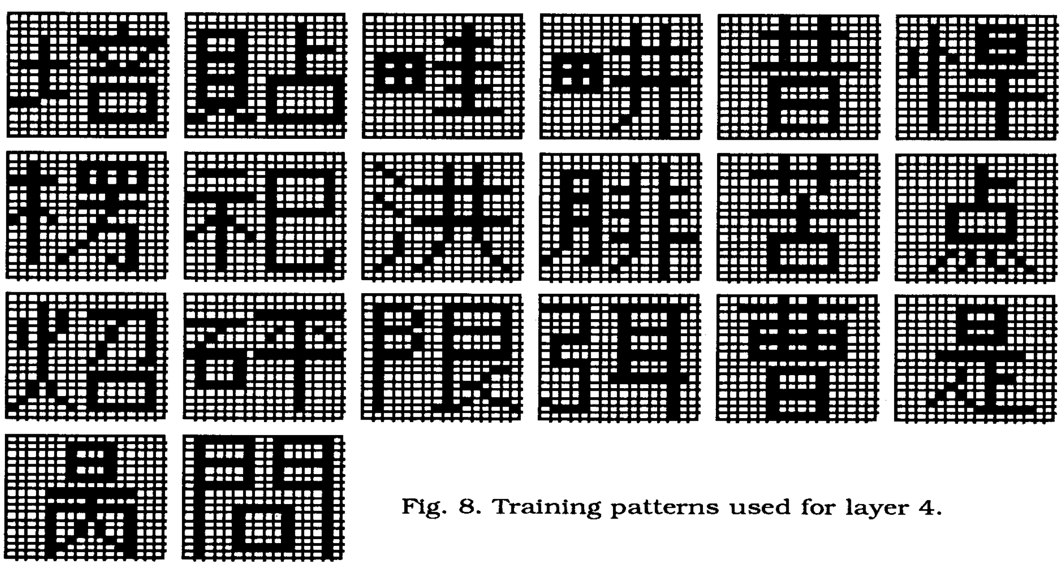

Fig. 8. Training patterns used for layer 4.